**Weekly Report**

**2019.0107-2019.0113**

1. **This Week**

**AI FOR VIS Project**

1. 与王老师、闽峰、浩哲讨论，梳理思路。
2. 方法思路回顾：
3. 无监督特征提取。
4. 人交互的检验、过滤这些特征（因子分析）
5. 基于这些特征做无监督聚类，约等于实现label
6. 通过神经网络在label后的数据集上做分类
7. 明确方法细节：
   1. beta-vae如何实现我们需要的因子分析：
8. 对图片集合X中的任意xi，可以通过训练后的vae模型得到p(z|xi)，对应于xi中的因子分布
9. 利用该分布对X进行聚类分析，研究各类中，各个因子对应的图像中的模式。实现方式：取各聚类中心的因子分布p(z|xc)，研究该类的典型模式，在p(z|xc)分布上采样得到潜变量向量z=[z1,z2,…,zn]（假设潜变量为n维），改变某个因子（zi）的值，得到新的潜变量向量z’，通过VAE模型中的decoder解码z与z’，观察解码后的生成图片，找到因子zi对应的模式。
10. 通过第（2）步，我们可以得到若干个图像中的模式，每个模式对应一个潜变量向量z（其中每个因子的取值范围应该在一个确定的区间段）。根据图片集中每个xi是否符合这些潜变量向量z的区间分布，完成对X的标注。
    1. 如何实现因子可视化分析：（插值）
11. 对于一个图片xi，求得分布p(z|xi)，基于p(z|xi)采样得到latent vector：z=[z1, z2, …, zn]
12. 研究第k个因子对应的模式（k属于[1,n]）。取向量zl=[z1, …, zkl, … , zn]，zr=[z1, …, zkr, … , zn]，其中zkl<zk<zkr.
13. 在zkl与zk，zk与zkr之间做差值得到一系列差值向量
14. 将zkl，zk，zkr与所有得到的差值向量通过decoder解码生成图片，并观察

备注：（a）由于n的值较大，可以通过取方差较小（方差前十）的因子进行分析。

（b）差值方法：线性/非线性差值（见Deep Feature Consistent Variational Autoencoder与Sampling Generative Networks）

（c）不一定只改变一个因子，可以同时改变几个因子

（d）latent vector是可以求均值，并做加减的，甚至还可以做相关性分析（计算后的attribute vector），还可以做投影

3.3 有人参与的因子分析及其可视化的工作量远比我之前想象的大。

1. 我们的方法属于半自动方法，已有的类似自动方法包括：（1）DeepCluster：等同于我们方法的1-4步，不同之处在于用CNN做特征提取与分类，聚类与分类迭代进行用于优化CNN参数；（2）Deep Embedded Cluster（DEC）：等同于我们方法的1-3步，使用SAE初始化，用深度神经网络聚类。

4.1 与闽峰讨论后，闽峰建议我们从全自动的开始，一步一个脚印的改代码，不然容易崩。所以本周我们先跑了DeepCluster模型（暂未出结果）。

4.2 根据我们自己的方法流程，我们计划下周开始跑VAE。并且开始撰写论文。

4.3 我自己与冯浩哲讨论了一下合作的可能性，希望下周跟您一起讨论一下。

**潮流项目**

1.后端实现与前端传输数据的接口（导入数据和数据修改模块已完成）。

2.之前说将潮流项目写成论文投电网技术，将该任务布置给了赵喆，本周给她写了论文撰写的提纲如下:

第一章 引言

* 什么是电网仿真计算？什么是潮流计算（电网仿真计算的一种）？潮流数据分析的重要性。
* 传统的潮流分析方法的局限性（依赖人工、分析局部等），于是需要引入可视分析。
* 潮流数据可视分析也有一部分工作，与之相比我们工作的不同之处或优势在于：1.已有工作大部分分析潮流结果，我们对计算过程进行了分析，面向如何调出最优潮流，支持潮流的调整 2.对潮流结果设计了新的可视化编码

第二章 相关工作

2.1 电网仿真计算及其分析

* 电网仿真计算有哪些，潮流是其中的一种。
* 电网仿真计算的分析方法（传统方法是什么，劣势在哪里）

2.2 电网数据可视分析

* 介绍已有的电网数据可视化工作

第三章 数据

* 数据规模（多少个节点与边）
* 数据的结构（抽象为网络结构，节点跟边上有哪些属性值，潮流的输入输出各是啥）

第四章 任务与系统总览

4.1任务分析

T1.观察潮流结果

T2.研究潮流收敛规律

T3.实现交互式的潮流调整过程

4.2系统总览

讲一下系统的结构结构，前后端如何衔接、分工

第五章 可视分析系统

5.1-5.X 讲一下各个视图中的visual encoding

5.(X+1) 交互

第六章 案例分析

第七章 讨论

第八章 总结

**南方电网项目**

1.与邵明、俊华等人讨论进展（邵明他们已经完成5篇实用性专利并交给了KH）。

2.治军在整理电力数据可视化论文，并构思综述框架。

3.与大家一起找了一些电能量数据的可视化文章，让治军、金旋、文杰一起阅读并整理一份报告，说明可做的电量问题与已有方法。

**其他**

1.给王老师整理了一份可考虑的，能将深度学习、机器学习方法应用的电网问题（见另一个pdf）。

**Working Hour: (except nap and eat time)**

9 小时/工作日

8 小时/周六

3 小时/周日

总工作时长：56小时。

**论文阅读**

**1. Analyzing the Effectiveness and Applicability of Co-training**

**2. beta-VAE: LEARNING BASIC VISUAL CONCEPTS WITH A CONSTRAINED VARIATIONAL FRAMEWORK**

**3. Sampling Generative Networks**

**4. Disentangling by Factorising**

**5. Deep Feature Consistent Variational Autoencoder**

**6. GANViz: A Visual Analytics Approach**

**to Understand the Adversarial Game**

1. **Progress**

|  |  |  |
| --- | --- | --- |
| Task | Progress | Time |
| AI For Vis | 1.确定方法细节 | 0226 |
| Power Flow Project | 1.前后端交接  2.撰写中文论文 | In one weeks  论文：放假前尽可能多写一些。 |
| Southern Power Grid Project | 1.解读数据、构思问题  2.构思综述框架  3.整理已有工作，发现问题 | - |
| SQC paper | Delayed | - |